Locally linear representation for subspace learning and clustering

Liangli Zhen, Zhang Yi, Xi Peng, Dezhong Peng

It is a key to construct a similarity graph in graph-oriented subspace learning and clustering. In a similarity graph, each vertex denotes a data point and the edge weight represents the similarity between two points. There are two popular schemes to construct a similarity graph, i.e., pairwise distance based scheme and linear representation based scheme. Most existing works have only involved one of the above schemes and suffered from some limitations. Specifically, pairwise distance based methods are sensitive to the noises and outliers compared with linear representation based methods. On the other hand, there is the possibility that linear representation based algorithms wrongly select inter-subspaces points to represent a point, which will degrade the performance. In this paper, we propose an algorithm, called Locally Linear Representation (LLR), which integrates pairwise distance with linear representation together to address the problems. The proposed algorithm can automatically encode each data point over a set of points that not only could denote the objective point with less residual error, but also are close to the point in Euclidean space. The experimental results show that our approach is promising in subspace learning and subspace clustering.

Introduction: As a useful technique in subspace learning and clustering methods, graph-oriented algorithm has become a hot topic in machine learning and computer vision, whose key is to construct a similarity graph [1]. In the graph, each vertex denotes a data point and the edge weight represents the similarity between the connected points. Currently, there are two ways to calculate the similarity among data points [2], i.e., Pairwise Distance based Scheme (PDS) and Linear Representation based Scheme (LRS). PDS measures the similarity by using the distance between two points, e.g., Laplacian Eigenmaps (LE) [3] adopts the Euclidean distance with heat kernel to build the similarity graph. Alternatively, LRS assumes that each data point could be denoted as a linear combination of some intra-subspace points [6]. Thus, the linear representation coefficients could be regarded as a kind of similarity. Recently, LRS has attracted a lot of interests from subspace learning and clustering community since it could be better to capture the real data structure. For example, Locally Linear Embedding (LLE) [4], Neighborhood preserving embedding(NPE) [5], Sparse Subspace Clustering (SSC) [6], and Low Rank Representation (LRR) [7].

It is notable that these algorithms have only involved one of the schemes, and suffered from some limitations. Particularly, pairwise distance based methods are sensitive to the noises and outliers compared with linear representation based methods, since the pairwise distance only depends on the two connected data points and ignores the global structure of the whole data set. On the other hand, the subspaces spanned by the data points are usually dependent. Therefore, it is probable that linear representation based algorithms will wrongly select inter-subspaces points to represent a point, and result in a degraded performance.

This letter proposes an algorithm, named as Locally Linear Representation (LLR), which incorporates pairwise distance into linear representation. It selects the points that not only could denote the objective point with less residual error, but also are close to the point in terms of Euclidean distance. Moreover, the solution of the proposed method can be analytic, which forms another advantage of LLR.

Locally Linear Representation (LLR): Our basic idea derives from a widely-accepted assumption in manifold learning that a topological manifold is a topological space that is locally homeomorphic to an Euclidean space. It implies that the intra-subspace points not only can linearly represent each other, but also are close to each other in terms of Euclidean distance. However, most previous works only focused on one aspect of them and accordingly developed linear representation based algorithms and pairwise distance based algorithms. As discussed in Section I, these methods encountered some problems.

Fig. 1 gives a toy example to illustrate the limitations of PDS and LRS and show the effectiveness of our approach. It is clear that either PDS or LRS will wrongly find the nearest neighbors for point A, while our approach will get the correct answer by selecting the points not only belong

(a) PDS   (b) LRS   (c) Our method
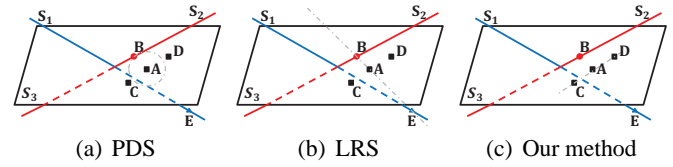
Fig. 1 *The geometric analysis of three different similarity graph construction strategies. There are three subspaces S_1, S_2, and S_3 lie in $\mathbb{R}^3$, where $dim(S_1) = 1$, $dim(S_2) = 1$, $dim(S_3) = 2$. Points A, C, D are draw from S_3, point B from S_2, and point E from S_1. Fig. 1(a) shows that the nearest neighbor of A is B in terms of Euclidean distance; Fig. 1(b) shows that the nearest neighbor of A is B and E in terms of linear representation based similarity; Fig. 1(c) shows that our method will select C and D as the nearest neighbor of A.*

to the same subspace with A, but also are close with A in terms of pairwise distance.

The objective function of our proposed approach is defined as following:

$$\min_{\mathbf{c}_i} \sum_{i=1}^{n} \lambda \|\mathbf{S}_i \mathbf{c}_i\| + (1-\lambda)\|\mathbf{x}_i - \mathbf{D}_i \mathbf{c}_i\| \quad s.t. \quad \mathbf{1}^T \mathbf{c}_i = 1, \quad (1)$$

where $\mathbf{c}_i \in \mathbb{R}^n$ is the similarity of a data point $\mathbf{x}_i \in \mathbb{R}^m$ over the corresponding dictionary $\mathbf{D}_i = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, \mathbf{0}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n]$, $\mathbf{S}_i$ is a diagonal matrix whose j^{th} diagonal element is the pairwise distance from $\mathbf{x}_i$ to the jth point in $\mathbf{D}_i$, $(j = 1, 2, \ldots, n)$, $\mathbf{1} \in \mathbb{R}^m$ is a vector composed with ones, and $\lambda \in [0, 1)$ is the trade off between pairwise distance strategy and linear representation strategy.

By solving the equation (1), it gives that

$$\mathbf{c}_i = \frac{\mathbf{M}_i^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{M}_i^{-1} \mathbf{1}}. \quad (2)$$

where $\mathbf{M}_i = \lambda \mathbf{S}_i^T \mathbf{S}_i + (1-\lambda)(\mathbf{x}_i \mathbf{1}^T - \mathbf{D}_i)^T (\mathbf{x}_i \mathbf{1}^T - \mathbf{D}_i)$.

However, the above solution is not sparse which produces too many connections inside the graph. It results in a high computing cost in subspace learning and clustering and decreases the discrimination between different subspaces. In this letter, we can get a more discriminate similarity matrix $\widetilde{\mathbf{C}}$ by preserving the k strongest connections for each data point as in [1]. In other words, we only keep k largest entries in $\mathbf{c}_i$, and set others to zeros. Consequently, we get a sparse similarity graph which has high discrimination.

Given a data set $\mathbf{X} \in \mathbb{R}^{m \times n}$, the proposed LLR algorithm constructs a similarity graph via the following steps:

(1) For each point $\mathbf{x}_i (i = 1, 2, \ldots, n)$, calculate its similarities, denoted as $\mathbf{c}_i$, to other points via (2).

(2) Get a more discriminated representation $\widetilde{\mathbf{c}}_i$ by keeping the k-strongest connected points to point $\mathbf{x}_i$ by using k-nearest neighbor searching on $\mathbf{c}_i$, and set all other elements in $\widetilde{\mathbf{c}}_i$ as zeros, where $i = 1, 2, \ldots, n$.

(3) Construct a similarity graph by treating each point as a vertex, and assigning the connection weight between the node i and j as $\mathbf{W}_{ij} = |\widetilde{\mathbf{C}_{ij}}| + |\widetilde{\mathbf{C}_{ji}}|$.

Once the similarity graph is built, multiple existing algorithms can be integrated with this similarity graph for subspace learning and subspace clustering. Subspace learning aims to find a projection matrix $\mathbf{P} \in \mathbb{R}^{m \times d}$ for transforming the high-dimensional datum into a low-dimensional space. Since the similarity graph is preserved in the transformation, following the embedding step of NPE [5], LLR graph can be used to develop a subspace learning algorithm. Then, the classification process can produce on the low-dimensional representation with much lower computational cost. Furthermore, subspace clustering refers to the problem of separating the data points according to their underlying subspaces [6]. The spectral clustering is the most popular algorithm for this task, but there is a challenge to measure the similarity between any two data points. We can use the the LLR graph as the similarity graph for spectral clustering. Similar to SSC [1, 6, 7], we can easily conduct spectral clustering via LLR graph, and achieve satisfactory performance, which will be shown by the simulation results presented in the next section.

Experimental verification: To verify the effectiveness of LLR in subspace learning and clustering tasks, we carried out the experiments over two

Table 1: Accuracy of 1-NN classifier for different subspace learning algorithms on AR database and Extended Yale Database B

Database	LLR	NPE	LPP
AR	**0.952**	0.734	0.881
ExYaleB	**0.953**	0.848	0.925

Table 2: Clustering quality (Accuracy (AC) and Normalized Mutual Information (NMI)) of spectral clustering algorithm based on the graph built by LLR, LRR, SSC, LLE, and NSC.

Database	Metric	LLR	LRR	SSC	LLE-Graph	NSC
AR	AC	**0.837**	0.771	0.767	0.396	0.361
	NMI	**0.929**	0.910	0.886	0.682	0.652
ExYaleB	AC	**0.883**	0.713	0.613	0.461	0.426
	NMI	**0.922**	0.772	0.684	0.540	0.539

popular databases, i.e., Extended Yale database B [8] and AR database [9]. The used Extended Yale database B contains 2014 near frontal face images of 38 individuals (58 images for each individual), where each image is downsized from 192×168 to 48×42. The used AR database contains 1400 face images without disguises distributed over 100 individuals (14 images for each subject), where each image is cropped from 165×120 to 55×40. For computational efficiency, we perform PCA to reduce the dimensionality of data by retaining 98% energy of the data. Moreover, to speed up LLR, we use 300 nearest neighbors as dictionary for each point. Similar to [1, 6, 7], in the all experiments, we reported the best result of the tested algorithms with tuned parameters.

We firstly investigate the performance of LLR in subspace learning by comparing with Neighborhood Preserving Embedding (NPE) [5] and Locality Preserving Projection (LPP) [10]. In the experiments, we randomly select a half of images from each subject (AR and Extended Yale B) for training purpose, and use the remaining as testing data. Table 1 reports the classification accuracy of k-Nearest Neighbor Classifier ($k = 1$, i.e., 1-NN) with the evaluated algorithms. From the results, we could see that the proposed algorithm outperforms the other methods with considerable performance margin. For example, the classification accuracy of LLR is at least higher 7.1% than the other tested methods over AR, and the figure is about 2.8% over Extended Yale Database B. Note that, NPE constructs a similarity graph based on linear representation, while LPP obtains the graph by computing the distance between two data points.

As discussed in the Section II, the graph built by LLR could be used for subspace clustering in the framework of spectral clustering. In this part, LLR was compared with several state-of-the-art algorithms. LRR [7], SSC [6], LLE-graph, and Normalized Spectral Clustering (NSC) [11]. Two popular metrics, Accuracy (AC) and Normalized Mutual Information (NMI) [1], are used to evaluate the clustering quality.

Table 2 reports the clustering results of LLR, from which we could have the following observations:

- LLR outperforms the other methods in AC and NMI with distinct advantage. For example, LLR is 6.6% and 1.9% higher than the second best method (LRR) over AR in AC and NMI, respectively. The figures are 17% and 15% over ExYaleB.
- LRR and SSC are two recently proposed algorithms, which are superior to LLE-graph and NSC which respectively are based on linear representation and pairwise distance (Euclidean distance + heat kernel).

Conclusion: Linear representation and pairwise distance are two popular ways to construct a similarity graph for subspace learning and clustering, which derive from one aspect of the assumption used in manifold learning such that they suffer from some limitations. To enhance the connectedness among intra-subspace points and disconnectedness among intra-subspace points, we proposed an algorithm to construct a similarity graph by enforcing the selected points are close not only in Euclidean distance but also could reconstruct the objective point with less residual error. Extensive experiments verified the effectiveness of our approach and corroborated the claims that combination of linear representation and pairwise distance are helpful to improve the discrimination of data.

References

1 Peng, X., Zhang, L., Zhang, Y.: 'Constructing L2-Graph For Subspace Learning and Segmentation', preprint arXiv:1209.0841v4, 2012.
2 Zhen, L., Peng, X., Peng, D.: 'Local Neighborhood Embedding for Unsupervised Nonlinear Dimension Reduction', Journal of Software, 2013, **8**, (2), pp. 410-417.
3 Belkin, M., Niyogi, P.: 'Laplacian eigenmaps for dimensionality reduction and data representation', *Neural computation*, 2003, **15**, (6), pp. 1373-1396.
4 S. Roweis, L. Saul: 'Nonlinear dimensionality reduction by locally linear embedding', *Science*, 2000, **290**, (5500), pp. 2323-2326.
5 X. He, D. Cai, S. Yan, H. Zhang: 'Neighborhood preserving embedding', IEEE Int. Conf. on Computer Vision, 2005, pp. 1208-1213.
6 Elhamifar, E., Rene V.: 'Sparse subspace clustering: Algorithm, theory, and applications', to appear in IEEE Trans. Pattern Anal. Mach. Intell., 2013.
7 Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: 'Robust recovery of subspace structures by low-rank representation', IEEE Transactions on Pattern Anal. and Mach. Intell., 2013, **35**, (1), pp. 171-184.
8 A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman: 'From few to many: Illumination cone models for face recognition under variable lighting and pose', IEEE Trans. Pattern Anal. Mach. Intell., 2001, **23**, (6), pp. 643-660.
9 A. Martinez, R. Benavente: 'The AR face database', CVC Tech. Report #24, 1998.
10 He, X., Niyogi P.: 'Locality Preserving Projections', Advances in Neural Information Processing Systems, 2003, **16**, pp. 153-160.
11 Ng, Andrew Y., Michael I. Jordan, Yair Weiss: 'On spectral clustering: Analysis and an algorithm', Advances in Neural Information Processing Systems , 2002, pp. 849-856.